IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Harbaugh

Application No.: 09/826690   Examiner: Smith, Traci

Date Filed: April 5, 2001   Group: 3629

For: METHOD FOR ADMITTING AN ADMISSIONS APPLICANT INTO AN
ACADEMIC INSTITUTION

AFFIDAVIT OF STEPHEN P. KLEIN

I, Stephen P. Klein, being duly sworn, hereby declare:

1. I hold a B.S. in Psychology from Tufts University, a M.S. in Industrial Psychology, with minors in Research Design and Statistics, from Purdue University, and a Ph.D. in Industrial Psychology, with minors in Labor Law and Tests and Measurements, from Purdue University.

2. Since 1974, I have served as a consultant to national and state licensing boards, including the National Conference of Bar Examiners, more than 25 state bar boards, the American Institute of Certified Public Accountants, the National Board of Professional Teaching Standards, the California Commission on Teaching, the Puerto Rico Board of Medical Examiners, and the Society of Actuaries.

3. Since 1975, I have worked at the RAND Corporation and currently hold the title of Senior Research Scientist. My work requires me to design and direct statistical studies in the fields of education, criminal justice, health, and military manpower. I have served on several National Research Council committees and have testified as a measurement and statistical expert before legislative committees.

4. I have expertise in the methods that are employed to obtain adequate inter-reader consistency in grading standards when the answers that were written by a large number of test takers to open-ended questions, including essay questions, are evaluated by different readers. This expertise includes knowledge of and experience with calibrated grading methods and procedures. I also have substantial expertise in the fields of statistics and research design.

{WP253999;1}

1

5. Calibrated grading is a well known term of art among individuals who participate in the administration and grading of examinations, such as essay exams, that require readers to evaluate the quality of the test-takers' responses. As used in this sense, "calibration" is the process graders go through to help assure that different graders will use consistent standards when evaluating answers. In a well calibrated grading process, a given answer will receive the same score regardless of which grader evaluated that answer's quality. It also will receive the same score if the same reader evaluated the answer at two or more different points in time.

6. Calibration is standard practice on large-scale tests that use multiple readers to evaluate test taker responses to written responses, such as on licensing exams for lawyers and teachers, and for the clinical or practical portions of licensing exams, such as for dentists. Calibration helps to assure that the evaluation of a test taker's performance is not a function of systematic differences among the graders who evaluate the responses. The calibration process is designed to reduce the likelihood that a test taker's rank, pass/fail status, or score, is a function of which evaluator assessed the quality of that test taker's answers or performance.

7. While calibration activities vary somewhat as a function of the specific features of the assessment program (such as the number of evaluators involved and the nature of responses being evaluated), they all include training in the criteria to be employed, round robin assessments of a common set of performances, and discussion and resolution of differences in judgment among the evaluators. That is why reader calibration, including round robin grading of a common batch of answers, is standard practice on large scale essay tests like the new SAT. Moreover, professional standards demand that calibration be used in high stakes testing programs where candidate performance is evaluated by readers, examiners, raters, or judges.

8. Calibration is needed regardless of whether an analytic or a holistic grading system is used. In an analytic system, the grader assigns points to various parts of a test taker's response, such as assigning a separate score for each of the several different issues in a case the test taker is expected to identify and discuss. In a holistic system, the evaluator assigns an overall score to the answer as a whole.

9. Calibration is needed regardless of whether scores are used for making norm-referenced decisions (i.e., deciding which test takers performed better than others) or for criterion-referenced decisions (i.e., deciding whether the test taker reached some specified level of proficiency). For example, scores on employment and college admissions tests are used to make norm-referenced decisions whereas scores on licensing exams are used to make criterion-referenced decision. The in-car portion of the driver's license test is a good example of test where the score is used to make a criterion-referenced decision.

10. In short, calibration helps to assure that the score assigned to a test taker's response to an open-ended question is based solely on the quality of that response rather than which reader, grader, rater, or judge evaluated that response.

11. I have no financial interest in the Harbaugh Patent Application (Application 09/826,690) or any related applications. I am not being paid for my services pertaining to the subject matter of this declaration
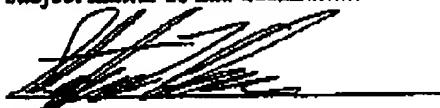
Stephen P. Klein

State of California
County of Los Angeles

I HEREBY CERTIFY that in this day, before me, an officer duly authorized to administer oaths, personally appeared Stephen P. Klein, who _____ is personally known to me or __✓__ has produced a driver's license as identification, was duly sworn to tell the truth, and having done so, says that the foregoing statements are true and correct.
IN WITNESS WHEREOF I set my signature and official seal this __23rd__ day of July 2005.
August

NOTARY PUBLIC

Jennifer Lin Prim Print Name

{WP253999;1}   3